PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 3 : C12P 21/00; C12N 15/00, 1/18 C12N 1/00

A1

(11) International Publication Number: WO 84/ 01173

(43) International Publication Date: 29 March 1984 (29.03.84)

(21) International Application Number: PCT/US83/01361

(22) International Filing Date: 9 September 1983 (09.09.83)

(31) Priority Application Numbers: 418,537
452,339

(32) Priority Dates: 15 September 1982 (15.09.82)
22 December 1982 (22.12.82)

(33) Priority Country: US

(71) Applicant: IMMUNO NUCLEAR CORPORATION [US/US]; 1951 Northwestern Avenue, Stillwater, MN 55082 (US).

(72) Inventor: MAHONEY, Walter, C. ; 7628 Dunmore Dr., Woodbury, MN 55125 (US).

(74) Agents: POPOVICH, Thomas, E. et al.; 2200 First Bank Place East, Minneapolis, MN 55402 (US).

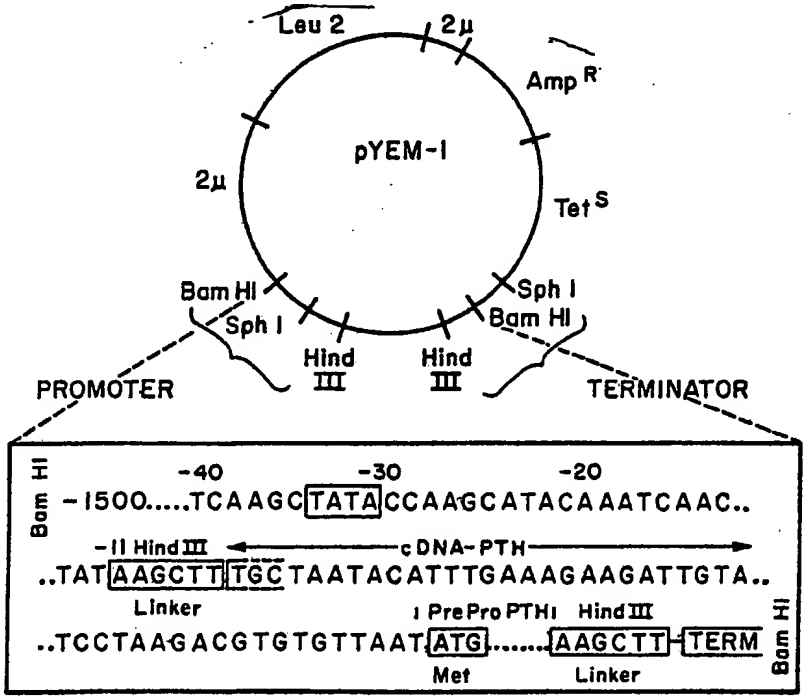
(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: PRODUCTION OF MATURE PROTEINS IN TRANSFORMED YEAST

(57) Abstract

A method for producing a mature protein in yeast transformed to express a corresponding precursor, wherein the mature protein sequence is contained in the precursor and is flanked proximally or both proximally and distally by a pair or triplet of basic amino acid residues. The method comprises proteolytic processing by an endopeptidase and exopeptidase present in the yeast. Yeast transformed by a plasmid containing a cDNA sequence encoding bovine preproparathyroid hormone is also disclosed.



PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 3 : C12P 21/00; C12N 15/00, 1/18 C12N 1/00 — A1

(11) International Publication Number: WO 84/ 01173

(43) International Publication Date: 29 March 1984 (29.03.84)

(21) International Application Number: PCT/US83/01361

(22) International Filing Date: 9 September 1983 (09.09.83)

(31) Priority Application Numbers: 418,537; 452,339

(32) Priority Dates: 15 September 1982 (15.09.82); 22 December 1982 (22.12.82)

(33) Priority Country: US

(71) Applicant: IMMUNO NUCLEAR CORPORATION [US/US]; 1951 Northwestern Avenue, Stillwater, MN 55082 (US).

(72) Inventor: MAHONEY, Walter, C. ; 7628 Dunmore Dr., Woodbury, MN 55125 (US).

(74) Agents: POPOVICH, Thomas, E. et al.; 2200 First Bank Place East, Minneapolis, MN 55402 (US).

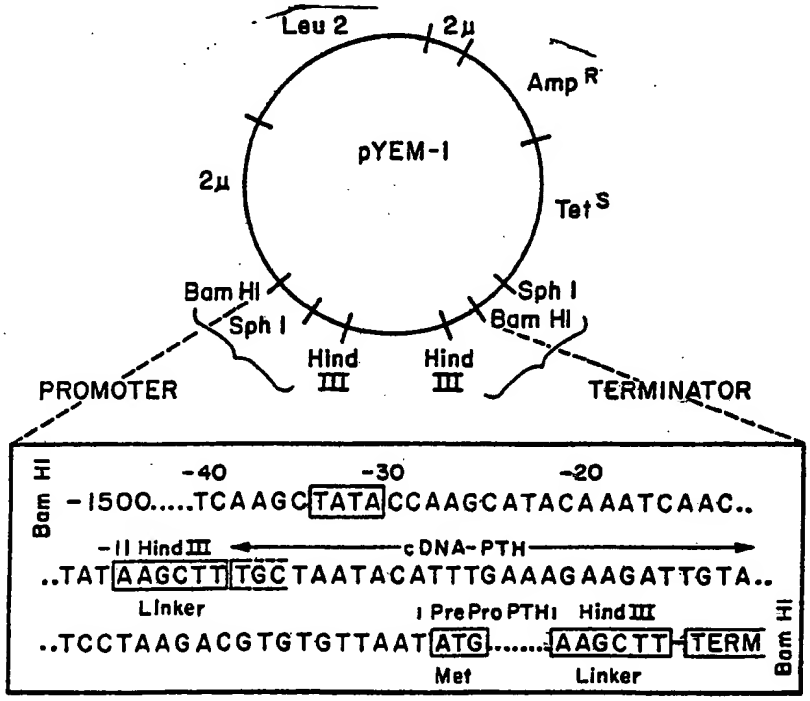
(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: PRODUCTION OF MATURE PROTEINS IN TRANSFORMED YEAST

(57) Abstract

A method for producing a mature protein in yeast transformed to express a corresponding precursor, wherein the mature protein sequence is contained in the precursor and is flanked proximally or both proximally and distally by a pair or triplet of basic amino acid residues. The method comprises proteolytic processing by an endopeptidase and exopeptidase present in the yeast. Yeast transformed by a plasmid containing a cDNA sequence encoding bovine preproparathyroid hormone is also disclosed.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	LI	Liechtenstein
AU	Australia	LK	Sri Lanka
BE	Belgium	LU	Luxembourg
BR	Brazil	MC	Monaco
CF	Central African Republic	MG	Madagascar
CG	Congo	MR	Mauritania
CH	Switzerland	MW	Malawi
CM	Cameroon	NL	Netherlands
DE	Germany, Federal Republic of	NO	Norway
DK	Denmark	RO	Romania
FI	Finland	SE	Sweden
FR	France	SN	Senegal
GA	Gabon	SU	Soviet Union
GB	United Kingdom	TD	Chad
HU	Hungary	TG	Togo
JP	Japan	US	United States of America
KP	Democratic People's Republic of Korea		

PRODUCTION OF MATURE PROTEINS IN TRANSFORMED YEAST

BACKGROUND OF THE INVENTION

1. Field of the Invention.

This invention relates to a method for producing a mature protein in transformed yeast and further relates to Saccharomyces cerevisiae transformed by a plasmid containing a preproparathyroid hormone cDNA insert.

2. Description of the Prior Art.

Recombinant DNA technology now makes it possible to isolate specific genes or portions thereof from higher organisms, such as man and other animals, and to transfer the genes or fragments to a microorganism species, such as E. coli or yeast. The transferred gene is replicated and propagated as the transformed microorganism may become endowed with the capacity to make whatever protein the gene or fragment encodes, whether it be an enzyme, a hormone, an antigen or an antibody, or a portion thereof. The microorganism passes on this capability to its progeny, so that in effect, the transfer results in a new strain, having the described capability.

Recombinant DNA conventionally utilizes transfer vectors. A transfer vector is a DNA molecule which contains genetic information which insures its own replication when transferred to a host microorganism strain. Plasmids are an example of a transfer vector commonly used in genetics. Although plasmids have been used as the

transfer vectors for the work described herein, it will be understood that other types of transfer vectors may be employed. Plasmid is the term applied to any autonomously replicating DNA unit which might be found in a microbial cell, other than the genome of the host cell itself. A plasmid is not usually genetically linked to the chromosome of the host cell. Plasmid DNA exists as doublestranded ring structures generally on the order of a few million daltons molecular weight, although some are greater than 10^8 daltons in molecular weight. They usually represent only a small percent of the total DNA of the cell. Transfer vector DNA is usually separable from host cell DNA by virtue of the great difference in size between them. Transfer vectors carry genetic information enabling them to replicate within the host cell.

Plasmid DNA exists as a closed ring. However, by appropriate techniques, the ring may be opened, a fragment of heterologous DNA inserted, and the ring reclosed, forming an enlarged molecule containing the inserted DNA segment.

Transfer is accomplished by a process known as transformation. During transformation, host cells mixed with plasmid DNA incorporate entire plasmid molecules into the cells. Once a cell has incorporated a plasmid, the latter is replicated within the cell and the plasmid replicas are distributed to the progeny cells when the cell divides.

Genetic information contained in the nucleotide sequence of the plasmid DNA, including heterologous DNA inserted into the plasmid, can in principle be expressed in the host cell. The inserted heterologous DNA typically representing

a single gene, is expressed when the protein product coded by the gene is synthesized by the organism.

Once a gene has been isolated, purified and inserted into a plasmid or other vector, the availability of the gene in substantial quantity is assured. After transfer of the vector into a suitable microorganism, the gene replicates as the microorganism proliferates. The vector containing the gene is easily purified from cultures of the host microorganism by known techniques and separable from the vector by restriction endonuclease cleavage followed by gel electrophoresis. The protein product expressed by the heterologous gene can also be recovered in substantial quantities from cultures of the host microorganism by harvesting the culture and retrieving the protein product contained in the harvested cells. (For futher detail of recombinant DNA technology, and an explicit exposition of the utility of producing proteins such as hormones, etc., by recombinant DNA technology, see U.S. Patent No. 4,237,224, issued December 2, 1980 to Cohen et al., and U.S. Patent No. 4,322,499, issued March 30, 1982 to Baxter et al. Patents and articles cited herein are incorporated by reference wherever such citations occur and shall be considered incorporated in their entirety as if set forth in full).

Recombinant DNA thus holds great promise for economically producing substantial quantities of useful proteins that are difficult or costly to isolate in such quantitites from mammalian tissue. A major and nearly universal problem in producing useful proteins, however, is the construction of the actual genetic material to be inserted into the transfer vector.

Conventional means provide for enzymatically preparing desired genetic material by reverse transcription. Mature messenger RNA (mRNA), which is chemically similar to DNA and retains most of the information coded in DNA, can be extracted from tissue in which the desired gene is active. mRNA is separated from other RNA material in the tissue and complementary DNA (cDNA) is produced by the enzyme reverse transcriptase, and at times polymerase 1 for the synthesis of the second strand. This cDNA, a complementary copy of mRNA and similarly containing the information coded in RNA, is often further altered in known ways to be suitable for insertion into a plasmid vector. (See W. Mahoney & S. Henikoff, Univ. of Washington Medicine, Vol. 8, No. 4, pp. 6-14 (Winter, 1981)).

cDNA enzymatically prepared by reverse transcription has the potential to express a protein chain identical to the protein expressed by tissue from which the mRNA was extracted. This alone is not sufficient, however, for the expression of desired mature animal proteins because many animal proteins, represented by such diverse classes as hormones, binding proteins, enzymes, antibodies, and collagen, are produced in nature in the form of larger precursors that are subsequently modified by cleavage to smaller bioactive forms commonly designated mature proteins. Thus, expression of cDNA synthesized by reverse transcription only has the potential to express the precursor of the mature protein product.

It has been known for several years that bacteria such as E. coli can remove the "pre" portion of its own secreated proteins. Examples include the processing of pre-ribose binding

protein, pre-galactose binding protein and pre-arabinose binding protein. (L. Randall, et al., Eur. J. Biochem., Vol. 92, pp. 411-415 (1978); L. Randall, S. Hardy, and L. Josefsson, Proc. Natl. Acad. Sci. USA, Vol. 75, pp. 1209-1212 (1978)).

S. Chan, et al., Proc. Natl. Acad. Sci. USA, Vol. 78, pp. 5401-5405 (1981) has exploited the ability of E. coli to remove the "pre" sequence. Chan, et al., modified cDNA for human preproinsulin to encode a hydrid "pre" sequence containing portions of E. coli and mammalian "pre" sequence. E. coli expressed the hydrid protein and correctly removed the "pre" sequence by intra-cellular processing. Thus, Chan, et al., was able to modify human preproinsulin cDNA in a way that would allow E. coli to produce proinsulin.

It is also known that yeast shares the ability to remove "pre" sequences from its own pre-proteins. Furthermore, when an E. coli preprotein was genetically engineered into yeast, pre-B-lactamase was processed to B-lactamase. (Roggenkamp, et al., Proc. Natl. Acad. Sci. USA, Vol. 78, No. 7, pp. 4466-4470 (1981)).

The above type of processing of preproteins, however, will not process to mature proteins many of the mammalian hormone precursors and many of the other interesting mammalian protein precursors in E. coli. These latter hormone and protein precursors contain a "pro" portion which is not processed by the enzymatic mechanism responsible for processing the "pre" portion of preproteins. As shown above, for example, the natural precursor for insulin, i.e. preproinsulin is processed in E. coli to form proinsulin.

Many investigators have been unable to express pre-proteins in yeast or E. coli, let

alone get processing. Expensive and time consuming, investigative efforts have focused almost exclusively on genetically eliminating the "pre" sequences and the "pro" sequences in attempting to express mature proteins without intermediates.

In several prior art approaches, the need for processing precursor proteins has been overcome. Insulin is the result of natural processing in human tissue involving cleaving two peptide chains, A and B, from the single large precursor preproinsulin and assembling the A and B chains to form the mature hormone insulin. The A and B chains are located within proinsulin and hence E. coli which processes preproinsulin to proinsulin does not produce the mature hormone insulin. An approach to obtaining mature insulin using E. coli employs chemically synthesized genes compatible with E. coli.

A double-stranded synthetic DNA-coding sequence for the insulin A chain was synthesized chemically from fundamental nucleotide units to yield the correct coding sequence. An extra amino acid (methionine) was added at one end. This end was fused to the bacterial gene for the enzyme B-galactosidase which results in accumulations of fused B-galactosidase-insulin-A-chain protein. This same procedure was repeated for the B-chain which resulted in the production of fused B-galactosidase-insulin-B-chain protein.

The fused proteins are insoluble in water and readily isolated from broken cells. The A and B chains of insulin are released from B-galactosidase at the extra methionine by cyanogen bromide cleavage and subsequently mixed together under conditions that allow formation of disulfide bonds between A

and B chains, yielding mature insulin. (W. Miller, J. of Pediatrics, Vol. 99, pp. 1-15 (1981); D. Goeddel, et al., Proc. Natl. Acad. Sci. USA, Vol 76, pp. 106-110 (1979)).

The above prior art approach overcomes the need for processing a precursor protein, but in turn requires processing of the fused B-galactosidase-insulin-A-chain and B-galactosidase-insulin-B-chain proteins to mature insulin. Moreover, chemical synthesis of the DNA coding sequences for A-chain and B-chain involves substantial costs, even when considering that the B-galactosidase-insulin-A-chain gene and B-galactosidase-insulin-B-chain gene after being synthesized are easily replicated for subsequent production of insulin. (D. Williams, et al., Science, Vol. 215, pp. 687-689 (Feb. 1982); W. Mahoney, Univ. of Wash. Medicine, supra).

The approach of chemically synthesizing DNA encoding for mature proteins has also been shown to be effective for bacterial production of human somatostation. (K. Itakura, et al., Science, Vol. 198, pp. 1056-1063 (1977)). However, insulin chains A and B and human somatostation are relatively small sequences and chemically synthesized DNA coded for them are relatively small. In the case of larger proteins, chemical synthesis of the DNA coding sequence coded for such proteins is prohibitively time consuming.

One prior art approach, now often followed, utilizes chemically synthesized DNA in conjunction with enzymatically prepared cDNA to produce a gene which instructs production of mature hormone in bacteria. Human growth hormone (HGH) is a protein of 191 amino acids, its precursor having an additional 26 amino acid "pre" portion. cDNA

encoding the precursor was enzymatically prepared from mRNA isolated from human pituitary tissue. The first useful cleavage site of the cDNA occurs at the site encoding amino acid residues 23-24 of HGH. Treatment of the cDNA with restriction endonuclease Hae III gives a DNA fragment of 551 base pairs which includes coding sequences for amino acids 24-191 of HGH. A gene fragment having coding sequences for residues 1-23 of HGH (and an initiation codon) was chemically synthesized. The two DNA fragments were combined to form a synthetic-natural hybrid gene which when inserted into a plasmid vector directed expression of mature HGH in E. coli. (D. Goeddel, et al., Nature, Vol. 281, pp. 544-548 (October 1979)).

Using a similar strategy of cleavage and reconstriction of DNA for the mature protein, R. Lawn et al., Nucleic Acids Research, Vol. 9, No. 22, pp. 6103-6114 (1981), expressed mature human albumin in E. coli.

This general approach, however, requires time consuming chemical synthesis of desired gene fragments, cleavage of cDNA assuming the availability of useful cleavage sites and difficult genetic construction of plasmids from DNA fragments. Furthermore, in both of the above examples, an initiator methionine was left at the NH_2 -terminal. The initiator methionines cannot practically be removed since HGH and albumin also have methionines located elsewhere in the sequence. Thus, removing the initiator methionine by cayanogen bromide cleavage, would result in cleavage at the other methionines. This would result in a protein split into cleaved fragments. Both the HGH and albumin produced by the above approach are "mature" proteins which start with methionines. Hence they

are not "real" mature proteins.

The prior art approaches set forth above illustrate that a major and nearly universal problem in producing mature proteins is the construction of the actual genetic material to be inserted into transfer vectors. Procedures exist for preparing cDNA from mRNA isolated from mammalian or other higher order animal tissue, but mammalian and higher order animal proteins are most often expressed as precursors and subsequently processed into the mature protein in cells of origin. The prior art has identified E. coli and yeast as microorganisms capable of processing precursors containing the "pre" portion, but this class of precursors excludes many of precursors of interest. The prior art thus has not identified a microorganism suitable for cloning mammalian and higher order animal genes which is capable of processing to mature proteins precursors of greatest interest. The prior art approaches attempt to solve the problem by constructing genes that code for mature protein. However, although procedures now exist for identifying nucleotide coding sequences for mature proteins, chemical synthesis of DNA sequences encoding mature proteins or fragments thereof for use in hydrid genes is costly and time consuming, often prohibitively so.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows inferred protein cleavage sites within the precursor of yeast α-factor, where "K" designates lysine and "R" designates arginine amino acid residues.

FIG. 2 shows the cDNA sequence encoding preproparathyroid hormone and the unique Pvu II and Hinf 1 cleavage sites.

FIG. 3 shows certain portions of the nucleotide sequence of the pYEM-1 plasmid.

SUMMARY OF THE INVENTION

In the present invention, a method is disclosed for producing protein in yeast transformed to express a corresponding precursor containing a pair or triplet of basic amino acid residues located proximally and/or distally adjacent to the protein portion of the precursor sequence comprising proteolytic processing by the yeast of the precursor at the site of such pairs or triplets of basic amino acid residues. The method comprises proteolytic processing by transformed yeast which contains an endopeptidase, designated herein as a trypsin-like enzyme or enzymes. The trypsin-like enzyme or enzymes proteolytically process the precursor at the site of such pairs or triplets of basic amino acid residues by cleaving at the distal side of such pairs or triplets. The method further comprises proteolytic processing by transformed yeast that contains an exopeptidase, designated herein a carboxypeptidase-B-like enzyme or enzymes. The carboxypeptidase-B-like enzyme or enzymes proteolytically process the precursor at the site of such pairs or triplets of basic amino acid residues by degrading such pairs or triplets of basic amino acid residues remaining distally adjacent to the protein portion of the precursor sequence after the cleavage by the trypsin-like enzyme or enzymes.

In the present invention, the above method is further disclosed for proteolytic processing of proto-proteins to mature proteins. Proto-proteins, defined with greater specificity infra, consist generally of precursor proteins in which the

protein portion of the precursor sequence is identical in structure to the mature protein except for the abscence of the amino terminal and the carboxyl terminal in the precursor sequence. The above method is also disclosed for proteolytic processing of certain non-proto-proteins. For example, the above method is disclosed for proteolytic processing of preproinsulin or proinsulin to mature insulin. The above method is disclosed for producing mammalian insulin generally as well as human, bovine, and porcine insulin specifically. According to the method, preprocalcitonin and procalcitonin may be proteolytically processed by transformed yeast to form mature calcitonin or a calcitonin relative in the case of animal calcitonin generally and human, bovine, and porcine calcitonin specifically.

In the present invention, a recombinant DNA plasmid transfer vector useful for transforming yeast comprising a DNA sequence comprising the preproparathyroid gene cDNA sequence is disclosed as well as the plasmid pYEM-1 and yeast transformed by a plasmid comprising the above transfer vector and yeast transformed by the plasmid pYEM-1.

DESCRIPTION OF THE SPECIFIC EMBODIMENT

Proto-proteins may consist of precursors for which DNA and mRNA encoding the precursors naturally occur in animals. This type of proto-protein is designated source natural proto-proteins. Proto-proteins may also consist of precursors in which synthetic DNA encodes the precursor. This type of proto-protein is designated source synthetic proto-protein. For example, by chemical synthesis, or alternatively by enzymatic cleavage, rearrangement and subsequent fusion, DNA can be synthesized so that the precursor which

it encodes has the cleavage properties discussed below. Production of mature protein might be enhanced by transforming yeast with synthetic DNA encoded for a precursor having repetitive sequences of the mature protein, each sequence being flanked by appropriate cleavage sites.

Source natural proto-proteins are illustrated by, but not limited to, certain hormone precursors, including preproparathyroid (J. Habener & J. Potts, The New England Journal of Medicine (Second Part), Vol. 299, No. 12, pp. 635-643 (Sept. 1978)), preprosomatostatin (P. Hobart, et al., Nature, Vol. 288, pp. 137-139 (November 1980)), AVP-NpII precursor to arginine vasopressin and its corresponding neurophysin (H. Land, et al., Nature, Vol. 295, pp. 299-303 (January 1982)), cortitropin B-lipotropin precursor to corticotopin (ACTH) and B-lipotropin (B-LPH) (S. Nakanishi, et al., Nature, Vol. 278, pp. 423-427 (March 1979)), preproglucagon (P. Lund, et al., Pro. Natl. Acad. Sci. USA, Vol. 79, pp. 345-349 (January 1982)), and pro-opiomelanocortin (POMC) precursor to B-endorphin and Met- and Leu-enkephalin precursor (M. Comb, et al., Nature, Vol. 295, pp. 663-666, (February 1982)).

Source natural proto-proteins are also illustrated by melittin precursor (G. Suchanek, et al., Eur. J. Biochemistry, Vol. 60, pp. 309-315 (1975); G. Suchanek, et al., Proc. Natl. Acad. Sci. USA, Vol. 75, pp. 701-704 (1978)) and serum albumin precursors (R. Lawn, et al., Nucleic Acids Research, Vol 9, No.22, pp. 6103-6114 (1981)).

As reported in the above citations, these precursors contain within their sequence at least one mature protein sequence. Where there is a single mature protein sequence contained in the precursor it is flanked proximally by a pair or

triplet of basic amino acid residues consisting of lysine and/or arginine and is flanked distally by either the carboxyl-terminal of the precursor or a pair or triplet of basic amino acid residues lysine and/or arginine. If there are several mature protein sequences contained in the precursor, at least one of the mature protein sequences is flanked proximally by a pair or triplet of such basic amino acid residues and is flanked distally by either the carboxyl-terminal of the precursor or a pair or triplet of such basic amino acid residues. Any precursor protein falling within this description is defined herein as a proto-protein, whether it be source natural or source synthetic.

As reported in the above citations in connection with observing the production of mature proteins in mammals and other higher order animals, the cleavage site located on the distal side of a pair or triplet of such basic amino acid residues is readily attacked by endopeptidases with trypsin-like activity. After endopeptidase cleavage, any residual basic residues remaining adjacent to and on the distal side of the mature protein are susceptible to degrading, i.e. selective removal, by exopeptidases with activity resembling that of carboxypeptidase-B.

Thus, for example, in preproparathyroid hormone the mature protein is flanked proximally by the basic triplet lysine-lysine-arginine and is flanked distally by the carboxyl-terminal of the precursor. A single cleavage by a trypsin-like enzyme is sufficient to produce the mature hormone. In other proteins such as the glucagon precursor, two mature glucagon proteins are flanked both proximally and distally by a basic pair lysine-arginine. Combined cleavage by a trypsin-like

enzyme and degradation of the resulting carboxyl-terminal by a carboxypeptidase-B-like enzyme are required to produce the mature proteins.

The method of the present invention comprises preteolytic processing by yeast of proto-proteins to mature proteins. In the method, transformed yeast naturally containing a trypsin-like enzyme or enzymes and a carboxypeptidase-B-like enzyme or enzymes, proteolytically release mature proteins from larger precursors. These enzymes will effectively cleave and degrade proto-proteins to mature proteins. This is confirmed by a trypsin-like cleavage, discussed infra, of preproparathyroid hormone yielding mature parathyroid hormone. This is further confirmed by yeast processing its own mating factor, α-factor. (T. Tanaka, et al., J. Biochemistry, Vol. 82, pp. 1681-1687 (1977)). As shown in FIG. 1, the nucleotide sequence of α-factor shows that yeast naturally expresses a precursor containing four distinct codings for mature α-factor. Three of the four α-factors in the precursor are flanked distally by a pair of basic amino acids residues. A trypsin-like cleavage in combination with a carboxypeptidase-B-like degrading naturally yields correctly processed C-termini for these three α-factors. After a trypsin-like cleavage, N-termini of the four α-factors are flanked proximally by a series of several glutamic acid and alanine amino acid residues. These latter residues are in turn removed by an aminopeptidase. The foregoing natural endopeptidase and exopeptidase activity in yeast in combination with the virtual uniform presence of pairs and triplets of lysine and/or arginine flanking mature hormone sequences in proto-proteins underlies the present invention.

Although preproinsulin and proinsulin containing disulfide bonds are not proto-proteins as defined herein they will nevertheless undergo proteolytic processing in yeast transformed to express the preproinsulin or proinsulin. A pair or triplet of basic amino acid residues are located distally and/or proximally adjacent to the insulin-A-chain and the insulin-B-chain portions of the sequence which constitute the protein portion of the precursor preproinsulin and proinsulin sequence. The requisite disulfide bonds between the insulin-A-chain portion of the sequence and the insulin-B-chain portion of the sequence will be formed in yeast. (cf. the numerous examples of disulfide bond formation in yeast disclosed in M. Dayhoff, Atlas of Protein Sequence and Structure, Vol. 5 and Supplements 1, 2 & 3 (National Biomedical Research Foundation, Georgetown University Medical Center, Washington, D.C. 20007 (1972, 1973, 1976, and 1981))). Proteolytic processing at the site of such pairs or triplets of basic amino acid residues will yield mature insulin from preproinsulin or proinsulin containing the disulfide bonds.

In the absence of disulfide bond formation between the insulin-A-chain portion of the sequence and the insulin-B-chain portion of the sequence, proteolytic processing will yield insulin-A-chain and insulin-B-chain, which may be caused in turn to attach to one another by disulfide bonds by conventional means to form mature insulin. In this case, the insulin-A-chain and insulin-B-chain may be considered mature proteins and preproinsulin and proinsulin without disulfide bonds may be considered a proto-protein according to the above discussion of proto-proteins.

Mature calcitonin contains disulfide bonds between the cysteines located at positions 1 and 7 of the sequence, contains a carbohydrate attached at the sequence at position 3, and the proline at position 32 has been amidated to pro-amide while the glycine at position 33 has been removed. Preprocalcitonin and procalcitonin will contain the requisite disulfide bonds. (cf. the numerous examples of disulfide bond formation in yeast as disclosed in Dayhoff, supra). A carbohydrate will be attached at position 3 in calcitonin. Preprocalcitonin and procalcitonin will undergo proteolytic processing in yeast transformed to express the preprocalcitonin or procalcitonin. A pair of basic amino residues are located proximally adjacent to the 33 amino acid sequence, while a triplet is located distally adjacent to the 33 amino acid sequence. It is expected that amidation of the proline located at 32 will occur in yeast after the cleavage distall to and degradation of the triplet. (cf. numerous examples of amidation in yeast as disclosed by Dayhoff, supra). In the event that a carbohydrate differing from the carbohydrate of mature calcitonin is formed by the yeast, the calcitonin relative containing the differing carbohydrate may be converted to mature calcitonin by conventional means. In the event that amidation following cleavage and degradation is suppressed, the calcitonin relative lacking the amidation may also be converted to mature calcitonin by conventional means.

By reverse transcription, cDNA can be prepared encoding any proto-protein of interest by isolating mRNA from tissues expressing the protein. Although many hormone and other protein genes have

already been cloned in E. coli, yeast has heretofore not been the host of choice. cDNA not previously cloned in yeast can be rendered compatible with a yeast host by proper codon selection (J. Bennetzen & B. Hall, J. Bio Chem., Vol. 257, pp. 3026 (1982)) and by site specific mutagenesis of the cDNA (G. Simmons, et al., Nucleic Acid Research, Vol. 10, pp. 821 (1982)).

Thus, one of the fundamental problems with producing useful mature proteins by recombinant DNA techniques has been simplified in the case of mature proteins derived from proto-proteins. cDNA, although readily available for most proteins by reverse transcription of mRNA isolated from animal tissue, will express the precursor of the mature protein. Yeast, but not E. coli, has the requisite enzymes to process expressed proto-proteins, preproinsulin, or proinsulin to mature protein or insulin.

EXPERIMENTAL

In order to demonstrate the present invention, the following experiment was carried out.

The plasmid YEp-13 was obtained from Dr. Steven Henekoff, Fred Hutchinston, Dept. of Developmental Biology, Seattle, Washington, and can be constructed according to J. Broach, et al., Gene, Vol. 8, pp. 121-133, (1979). The gene which encodes yeast alcohol dehydrogenase 1 was modified according to Hitzelman, et al., Nature (London), Vol. 293, pp. 717-722 (1981), allowing the isolation of the transcription signals. These sequences, including the cloning site, were provided by Dr. G. Ammera. The plasmid YEp-13 was modified so that the tet^R gene of YEp-13 was interrupted at the Bam Hl site with the yeast alcohol dehydrogenase 1 gene promotor and RNA polymerase

stop sequences. A Hind III site between the latter two elements provided the cloning site. These modifications of plasmid YEp-13 were accomplished by methods set forth generally in U.S. Patent 4,237,224, supra, and Methods of Enzymology, Vols. 65 and 68 wherein such methods are reviewed.

The cDNA sequence coding bovine preproparathyroid hormone, shown in FIG. 2 and further described in B. Kemper, et al., Hormonal Control of Calcium Metabolism (Ed. by D. Cohn, et al., published Excerpta Medica at Amsterdam, Oxford, and Princeton 1981) at pp. 19, was obtained from Dr. Byron Kemper, Department of Physiology and Biophysics and School of Basic Medical Sciences, University of Illinois-Urbana. This cDNA sequence was restricted with the enzymes PVU II and Hinf 1 at the sites shown in FIG.2. These enzymes were obtained from New England Biolaboratories, Beverly MA. The Hinf 1 site shown in FIG. 2 was filled with nucleotides using the enzyme DNA polymerase 1 (the large fragment) which was obtained from New England Nuclear, Boston, MA. This modified sequence was then blunt-end ligated to Hind III linkers and restricted with the enzyme Hind III. The Hind III linkers and Hind III enzyme were obtained from New England Biolaboratories, supra. The resulting DNA fragment was then ligated into the Hind III site of the modified plasmid YEp-13 forming a novel plasmid. This plasmid was designated pYEM-1. FIG. 3 shows certain portions of the nucleotide sequence of pYEM-1. The foregoing construction of pYEM-1 was accomplished by methods set forth generally in U.S. Patent No. 4,237,224, supra, the BLR M13 handbook, and Methods of Enzymology, Vols. 65 and 68 wherein such methods are

reviewed.

After constructing pYEM-1, yeast cells were transformed with the plasmid using the methods of Beggs, Nature (London), Vol. 275, pp. 104-109 (1978) and Hinnen, et al., Proc. Natl. Acad. of Sci. USA, Vol. 75, pp. 1929-1933 (1978). Because pYEM-1 has the yeast leu 2 gene, the use of a leu 2 negative strain of yeast was used in the transformation for the purposes of selecting successful transformants. Yeast strain, X1069-2D, a strain of Saccharomyces cerevisiae defective in leu-2 function, was obtained from the Yeast Genetic Stock Center, Univ. of Califorinia-Berekeley.

Of course any other defective yeast strain, including strains within Saccharomyces pombe and other species, could be used. All that is required is that a complementation system be established between the yeast strain and the cloning/expression vector and that the vector be stabily maintained in yeast. For example, a Trp 1 strain could be used if the Trp 1 gene was on the vector. To date, several stable transformation systems have been described. (A. Hinnen and B. Meyhack, Current Topics in Microbiology and Immunology, Vol. 96, pp. 101-117 (1981); C. Hollenberg, Current Topics in Microbiology and Immunology, Vol. 96, pp. 119-144 (1981)).

The transformed yeast cells containing plasmid pYEM-1 were grown in a leucine deficient media containing 5% glucose, yeast extract, yeast nitrogen base and other nutrients suitable for yeast strain X1069-2D. After 24 hours of growth at 30°C, the media was collected and the yeast cells lysed. Bioassay was performed according to conventional techniques and PTH

radioimmunoassay was performed using Immuno Nuclear Corporation (Stillwater, MN) assays specific to the N-terminal, mid-molecule, and C-terminal regions of parathyroid hormone. The following table shows that both immunologically cross-reactive parathyroid hormone and biologically active parathyroid hormone is being produced in yeast.

TABLE

	PTH N-terminal RIA*	PTH Mid-molecule region RIA*	PTH C-terminal RIA*	Bioassay*
Cell lysate				
pYEM-1	16	16	16	10
control	0	0	0	0
Media				
pYEM-1	2	2	2	0.015
control	0	0	0	0

*expressed in nanomoles/ml

To confirm that correct processing had occurred, 50 ml of culture was prepared in which the parathyroid hormone producing yeast were grown in media containing ^{35}S methionine (80 μ ci/ml). After an overnight growth the cells were removed by centrifugation. The media was then incubated with specific N-terminal parathyroid hormone antibody. After two hours the antibody-antigen complex was recovered by centrifugation and washed three times with new media followed by an ether wash. This complex contained about 7,000 cpm of ^{35}S methionine incorporated into protein after TCA precipitation. This mixture was applied to a Beckman 890D

sequencer according to the methods of Mahoney and Nute, Biochem. Vol. 19, pp. 4436 (1980) and subsequently degraded 40 cycles. Sequence analysis demonstrated that the ^{35}S methionine was all contained in cycles number 8 and 18. In mature PTH, methionine appears only at positions 8 and 18 in the sequence. If preproparathyroid hormone expressed by the yeast was left unprocessed, we would expect ^{35}S methionine in cycles 1, 2, 7, 11, 14, 49, and 59 reflecting the appearance of methionine at positions -31, -30, -25, -21, -18, +8, +18 in the preproparathyroid sequence.

The novel microorganism yeast strain X1069-2D transformed by novel plasmid pYEM-1, designated X1069-2D-pYEM-1, was placed on permanent deposit in the Northern Regional Research Center, U.S. Dept. of Agriculture, Peoria, Illinois 61604 on September 8, 1982. The NRRL number for X1069-2D-pYEM-1 is Y-15153. The plasmid pYEM-1 and the transfer vector contained therein may be removed from this novel yeast strain by known means.

While the invention has been described in connection with a specific embodiment thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations uses, or adaptations of the invention within the scope of the appended claims.

CLAIMS

1. A method for producing protein in yeast transformed to express a corresponding precursor containing a pair or triplet of basic amino acid residues located proximally and/or distally adjacent to the protein portion of the precursor sequence, comprising proteolytic processing by the yeast of the precursor at the site of such pairs or triplets of basic amino acid residues.

2. The method of claim 1 wherein the proteolytic processing by yeast of the precursor at the site of such pairs or triplets of basic amino acid residues comprises cleaving, by a trypsin-like enzyme or enzymes present in the transformed yeast, at the distal side of such pairs or triplets of basic amino acid residues.

3. The method of claim 2 wherein the proteolytic processing by yeast of the precursor at the site of such pairs or triplets of basic amino acid residues further comprises degrading, by a carboxy-peptidase-B-like enzyme or enzymes present in the transformed yeast, of any such pairs or triplets of basic amino assay residues remaining distally adjacent to the protein portion of the precursor sequence after the cleavage by the trypsin-like enzyme or enzymes.

4. The method of claim 1 wherein the corresponding precursor is a proto-protein.

5. The method of claim 4 wherein the proto-protein is source synthetic proto-protein.

6. The method of claim 4 wherein the proto-protein is source natural proto-protein.

7. The method of claim 6 wherein the source natural proto-protein is bovine preproparathyroid hormone.

8. The method of claim 1 wherein the protein is mammalian insulin and the corresponding precursor is mammalian preproinsulin or proinsulin.

9. The method of claim 8 wherein the mammalian insulin and mammalian preproinsulin or mammalian proinsulin are members respectively of the group consisting of human insulin and human preproinsulin or human proinsulin, bovine insulin and bovine preproinsulin or bovine proinsulin, and porcine insulin and porcine preproinsulin or porcine proinsulin.

10. The method of claim 1 wherein the protein is animal calcitonin or an animal calcitonin relative and the precursor is animal preprocalcitonin or animal procalcitonin.

11. The method of claim 10 wherein the animal calcitonin or animal calcitonin relative and the animal preprocalcitonin or animal procalcitonin are members respectively of the group consisting of human calcitonin or human calcitonin relative and human preprocalcitonin or human procalcitonin, bovine calcitonin or bovine calcitonin relative and bovine preprocalcitonin or bovine procalcitonin, and porcine calcitonin or porcine calcitonin relative and porcine preprocalcitonin or pro-calcitonin.

12. The method of claim 1 wherein the yeast is Saccharomyces cerevisiae or Saccharomyces pombe.

13. The method of claim 12 wherein the yeast is Saccharomyces cerevisiae.

14. A recombinant DNA plasmid transfer vector useful for transforming yeast comprising a DNA sequence comprising the preproparathyroid gene cDNA sequence.

15. The plasmid pYEM-1

16. Yeast transformed by a plasmid comprising the transfer vector of claim 14.

17. Yeast transformed by the plasmid of claim 15.

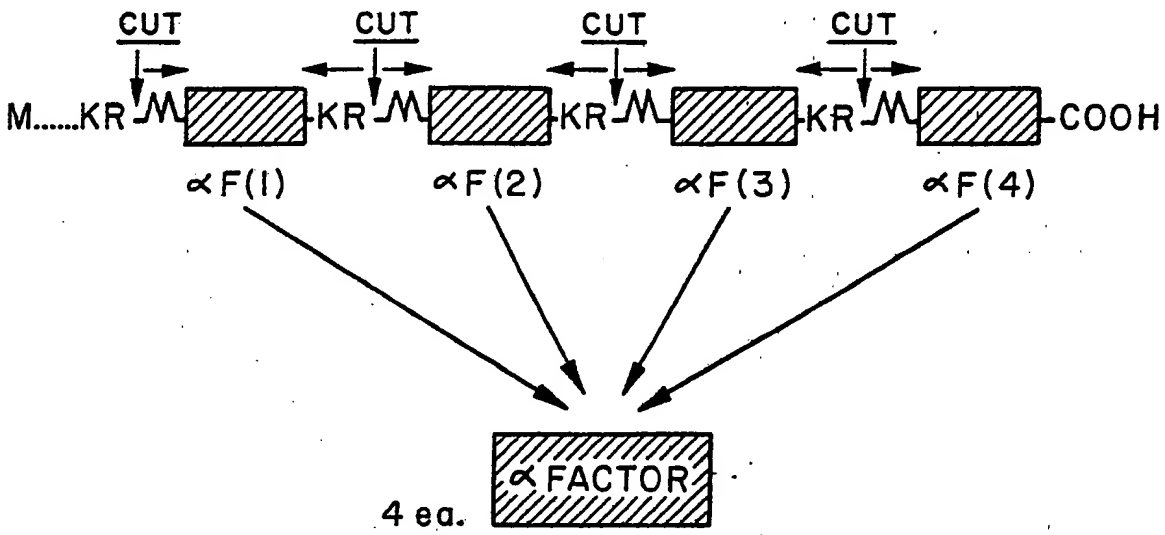


Fig 1

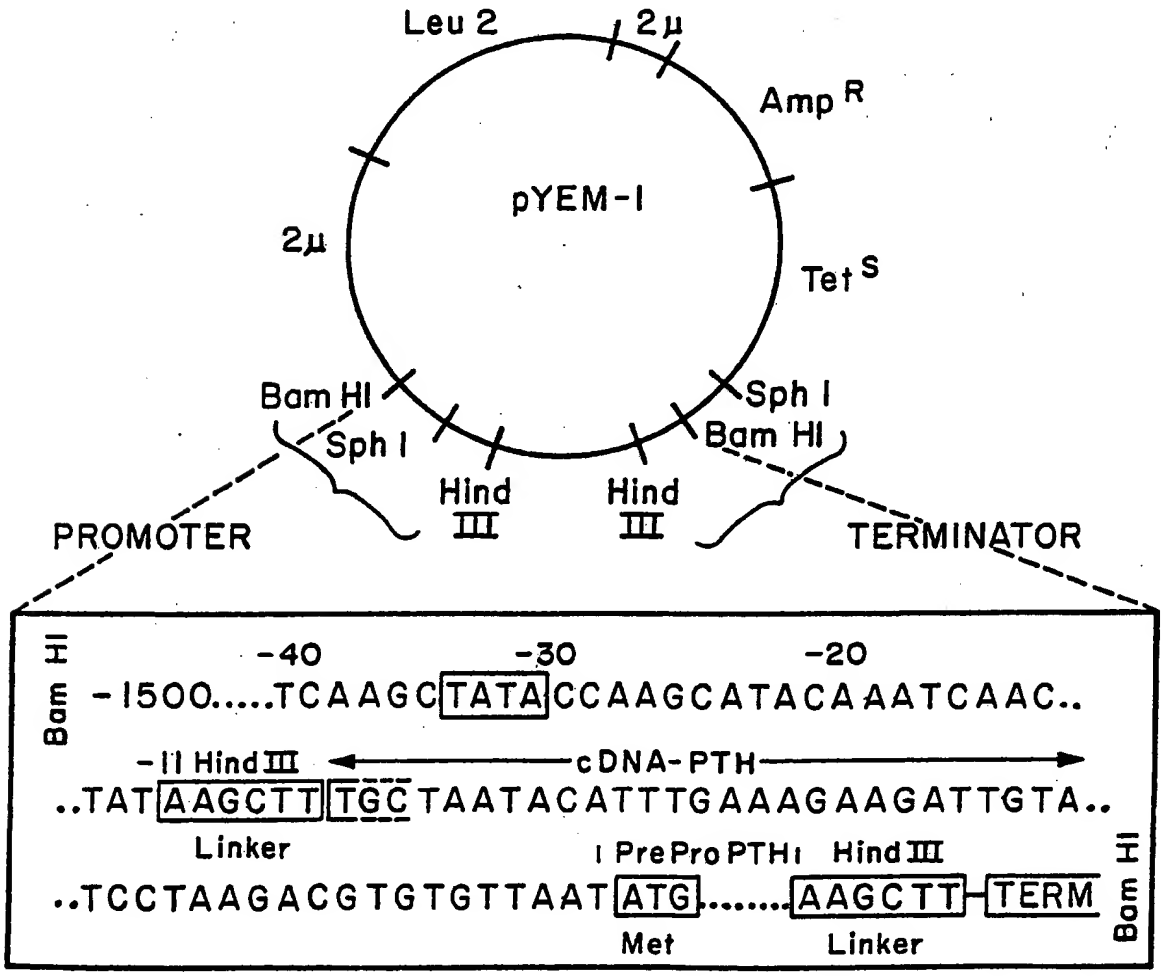


Fig 3

PvuII Cleavage Site C/T

```
5' G GGG GGG GGG GGG GGG GGG GTT TTA TCA GCC TTC TCA GGT TTA CTC AAC TTT GAG AAA GCA TCA GCT GCT AAT ACA TTT
               10          20          30          40          50          60          70

                                                  met met ser ala lys asp met val lys val met ile val met leu
GAA AGA AGA TTG TAT CCT AAG ACG TGT GTT AAT ATG ATG TCT GCA AAA GAC ATG GTT AAG GTA ATG ATT GTC ATG CTT
80          90          100         110         120         130         140         150

ala ile cys phe leu ala arg ser asp gly lys ser val lys lys arg ala val ser glu ile gln phe met his asn leu
GCC ATC TGT TTT CTT GCA AGA TCA GAT GGG AAG TCT GTT AAG AAG AGA GCT GTG AGT GAA ATA CAG TTT ATG CAT AAC CTG
 160         170         180         190         200         210         220         230

gly lys his leu ser ser met gly arg val glu trp leu arg lys lys leu gln asp val his asn phe val ala leu gly
GGC AAA CAT CTG AGC TCC ATG GAA AGA GTG GAA TGG CTG CGG AAA AAG CTA CAG GAT GTG CAC AAC TTT GTT GCC CTT GGA
 240         250         260         270         280         290         300         310

ala ser ile ala tyr arg asp gly ser ser gln arg pro arg lys lys glu asp asn val leu val glu ser his gln
GCT TCT ATA GCT TAC AGA GAT GGT AGT TCC CAG AGA CCT CGA AAA AAG GAA GAC AAT GTC CTG GTT GAG AGC CAT CAG
320         330         340         350         360         370         380         390

lys ser leu gly glu ala asp lys ala asp val asp val leu ile lys ala lys pro gln stop
AAA AGT CTT GGA GAA GCA GAC AAA GCT GAT GTG GAT GTA TTA ATT AAA GCT AAA CCC CAG TGA AAA CAG ATA TGA TCA GAT
 400         410         420         430         440         450         460         470

CAC TGT TCT AGA CAG CAT AGG GCA ACA ATA TTA CAT GCT GCT AAT GTG TTC ACC TTC TAT TAA GTG CCA GTA GTT CTA TGA
 480         490         500         510         520         530         540         550

                                                  Hinf Cleavage Site G/A

CCA ACC TTT ATT GCT AGC TGT GAT ACC TAC AAT TTT AAT TGA GTA TTT TGA TTC TAC TTT ATT CAT CTA AGA GCT CTT
560         570         580         590         600         610         620         630

TTA ATA ATT CTA TTT CTA TTG ATT CCA AAT AAA TGA AGT TAA GTA TTA AAA AAA AAA AAA AAA AAA AAA AAA AAA AAA
 640         650         660         670         680         690         700         710

AAA AAA AAA AAA AAA AAA CCC CCC CCC CCC CCC CCC CCC CCC CCC CCC 3'
 720         730         740         750         760
```

Fig 2

INTERNATIONAL SEARCH REPORT

International Application No PCT/US83/01361

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [3]

According to International Patent Classification (IPC) or to both National Classification and IPC

INT. CL.[3] C12P21/00,C12N15/00,1/18,1/00

U.S. CL. 435/68,172,256,317

II. FIELDS SEARCHED

Minimum Documentation Searched [4]

Classification System	Classification Symbols
U.S.	435/68,172,256,317

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [5]

CHEMICAL ABSTRACTS FILES 308,309,310,320 and 311
BIOSIS FILES 5,55 and 255

III. DOCUMENTS CONSIDERED TO BE RELEVANT [14]

Category *	Citation of Document, [16] with indication, where appropriate, of the relevant passages [17]	Relevant to Claim No. [18]
Y	N, WILLIAMSON, GENETIC ENGINEERING 4 ACADEMIC PRESS PP108-125,1983.	1-17
A	N, WALTON, RECOMBINANT DNA ELSEVIER SCIENTIFIC PUBLISHING CO., PP.185-197 and 213-227.	1-17
A	GB, A 2068969 A, PUBLISHED 19 AUGUST 1981.	1-17

* Special categories of cited documents: [15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search [2]	Date of Mailing of this International Search Report [2]
25 NOVEMBER 1983	06 DEC. 1983
International Searching Authority [1]	**Signature of Authorized Officer [20]**
ISA/US	ALVIN E. TANENHOLTZ

Form PCT/ISA/210 (second sheet) (October 1981)